Ministry of Education of the Republic of Belarus
Educational institution
Belarusian State University of
Informatics and Radioelectronics

УДК 004.93’11

YANG
Xue Ying

RECOGNITION SYSTEM OF HUMAN PHYSICAL ACTIVITY BASED ON TEMPORAL FEATURES OF SENSOR DATA

Abstract
for a Master’s Degree
in the Specialty 1-45 80 01 Infocommunication Systems and Networks

Anatoly
PHD, Full professor
Boryskevich Anatoly

Minsk 2022

INTRODUCTION

Human activity recognition (HAR) is a technology that uses a variety of sensors to identify and classify its activity data. At present, the research on human activity recognition is widespread and important in the fields of medical treatment, daily health care, pension and so on. The traditional process of human activity recognition mainly includes data acquisition, data preprocessing, data segmentation, feature extraction and selection, model training and testing, etc. This is essentially machine learning for pattern recognition, using smartphone data to distinguish between different activities, training the data, and finally building and authenticating a model based on a specific feature.

By analyzing the advantages and disadvantages of wearable devices and smart phones in user activity recognition, this master thesis proposes a perception framework based on wearable devices and smart phones, and develops a data acquisition tool for training and testing data on wearable devices and smart phones. The accelerometer and gyroscope built into the smartphone can be used to obtain the original signal, set the sampling rate to 50HZ, and select 2.56 seconds and 50% overlap. Sliding window (128 readings/Windows). We preprocess, segment division, feature extraction and other operations on the original data collected by sensors to obtain descriptions of different types of data.

Data acquisition module, which collects real human activity data through accelerometer and gyroscope sensor embedded in smart phone. Set the sampling rate to 50HZ. In the data acquisition experiment, the normal life scene mode is simulated and simple human activities are carried out. In the pre-processing of data, normalized operation is used to process the collected data. In this master thesis, high-Chebyshev filter is designed because of the influence of gravity in human actual activities. The instrument is used to remove the effects of gravity.

Human activity data collected by sensors is a continuous form of data flow. There is no obvious interval segmentation between human actions, and the exact boundary of activities is difficult to define, which makes it difficult to carry out feature extraction in activity recognition. Therefore, it is necessary to divide the approximately infinite length of data stream into finite and short segments through segment segmentation. The segmentation method adopted in this master thesis is sliding window. This technology refers to the segmentation of time series by moving a fixed-length window along the time series and extracting the data in the window.

Feature extraction is an important part of the framework of human activity recognition, which abstracts the important information that can reflect and distinguish the category of objects by adopting the method of feature extraction in

time domain. After feature extraction and data preprocessing, the data set is divided into predictive recognition for training set and test set activity recognition model. The model first trains on training samples and then evaluates the performance of the model based on data from validation samples. Activity recognition features consist of time domain features, including variance and mean.

At the same time, random forest classifier was used for activity identification. Firstly, Bootstrap sampling was used to extract K samples from the original training set, and the sample size of each sample was required to be consistent with that of the original training set. Secondly, a decision tree model was constructed for each sub-sample, and K decision trees were trained. Finally, K decision trees vote for final classification. During the training process, the test set is used to carry out the 10 fold cross test, adjust the training parameters, find the training model with the highest prediction accuracy, and finally use the model to predict the test samples. In order to evaluate the performance of the system, four parameters TPR, FDR, PPV and FNR were selected for evaluation.

Four algorithms were compared: the K-Nearest Neighbor algorithm (KNN), Naive Bayes algorithm (NB), the support vector machine (SVM) and the random forest algorithm (RF). The classification accuracy was used as the algorithm performance index to judge the recognition accuracy. According to the current experimental results, among the four typical statistical learning algorithms, the random forest algorithm has the best performance and the recognition accuracy of human activities reaches 96.8%.

The aim of this master thesis is using smartphone sensor data to identify physical human activities, it was eventually able to accurately identify six physical human activities. To achieve this aim, the following tasks were solved in the dissertation: Data acquisition using mobile phone sensor; the data set is preprocessed; Feature extraction is carried out using time domain features; Human activity recognition using random forest classifier; Using Confusion matrix to identify the performance of human physical activities.

GENERAL DESCRIPTION OF WORK

Relevance of the subject

The work corresponds to paragraph 1 «Digital information and communication and interdisciplinary technologies based on them production artificial intelligence and robotics» of the State Program of innovative development of the Republic of Belarus for 2021–2025. The work was carried out in the educational institution Belarusian State University of Informatics and Radioelectronics.

The aim and tasks of the work

The aim of the work is using smartphone sensor data to identify physical human activities, it was eventually able to accurately identify six physical human activities.

To achieve this aim, the following tasks were solved in the dissertation:

1 Data acquisition using mobile phone sensor.

2 The data set is preprocessed.

3 Feature extraction is carried out using time domain features.

4 Human activity recognition using random forest classifier.

5 Using Confusion matrix to identify the performance of human physical activities.

Personal contribution of the author

The content of the dissertation reflects the personal contribution of the author.

1 preprocessing ACC sensor data for improving extraction of time features;

2 Choice of the set of the time features for improving recognizing human physical activities;

3 Program realization of the random forest classifier and its performance evaluation for effective recognition of human physical activities;

Task setting and discussion of the results were carried out together with the supervisor Doctor of Technical Sciences, Full professor Boryskevich Anatoly

Testing and implementation of results

The main provisions and results of the dissertation work were reported and discussed at:

1 Xueying, Y.Human activity recognition based on random forest / Y. Xueying // Proc. of International Scientific and Technical Seminar, Technologies for the transmission and processing of information, Minsk. –2022. – P.77–81.

2 Xueying,Y.Human activity recognition based on Adaboost ensemble classifier / Y.Xueying, I.Boriskevich// Proc. of International Scientific and Technical Seminar ,Technologies for the transmission and processing of information, Minsk. –2022. – P.25–30.

3 Yang, Xueying. Recognition of human activities based on decision optimization model / Yang, Xueying, Huang Gang // Proc. of 2021 4th International Conference on Algorithms, Computing and Artificial Intelligence. – 2021. – P.1–8.

4 Yang, Xueying.Coordinated path planning for Multi-UAVs based on critical track points / Yang, Xueying,Huang Gang // Proc. of 2021 2nd International

Conference on Control, Robotics and Intelligent System. – 2021. – P.48–53.

The results of the thesis are used in scientific and technical products of Belarusian State University of Informatics and Radioelectronics.

Author's publications

According to the results of the research presented in the dissertation, 4 author's works was published, including: 4 articles in scientific journals recommended by the Higher Attestation Commission, with a total amount of 15 author's pages;

Structure and size of the work

The dissertation work consists of introduction, general description of the work, three chapters with conclusions for each chapter, conclusion.

The total amount of the thesis is 79 pages, of which 52 pages of text, 55 figures on 12 pages, 10 tables on 3 pages, a list of used bibliographic sources (27 titles on 2 pages), a list of the author's publications on the subject of the thesis (_4_ titles on 1 pages), graphic material on 9 pages.

Plagiarism

An examination of the dissertation « Recognition system of human physical activity based on temporal features of sensor data» by Yang xueying was carried out for the correctness of the use of borrowed materials using the network resource «Antiplagiat» (access address: https://antiplagiat.ru) in the online mode 31.03.2022. As a result of the verification, the correctness of the use of borrowed materials was established (the originality of the thesis is 97.3 %)

SUMMARY OF WORK

The **introduction** addresses the problems of the definition about human activity recognition and the basic structure of human activity recognition. Firstly, the definition of human activity recognition technology and the application field of human activity recognition are explained. Then, explained the process of human activity recognition: data acquisition, data pretreatment, data segmentation, feature extraction and testing in detail.

The **general description of work**: According to the framework of the human activity recognition system, in order to realize the use of smart phone sensor data to identify human body activities, and finally can accurately identify six kinds of human activities.

In this master thesis, five tasks of work are completed .Data collection using

mobile phone embedded sensor to simulate human normal life scenes. Since the actual human activities are greatly affected by gravity, chebyshev high-pass filter is used to preprocess the data set. The mean and standard deviation of time and frequency domain were used to extract features and principal component analysis was used to reduce the original feature scale. Using random forest classifier to recognize human activities can greatly improve the accuracy of human activity recognition. Four parameters TPR, FNR, PPV and FNR were introduced to evaluate the random forest classifier by using the confusion matrix

In the first chapter divided into six parts to describe. The first section introduces the application fields of human activity recognition, and summarizes the research status in the order of data acquisition, data preprocessing, data segmentation, and feature extraction and classifier evaluation. The second section analyzes the basic comparison between human activity recognition technology based on wearable devices and human activity recognition technology based on mobile phones, and discusses the advantages and disadvantages of these methods in detail. The third section introduces the application fields of human activity recognition, including intelligent medical treatment, monitoring system, intelligent home system, automatic driving system and intelligent prevention and control system. The fourth section describes the steps and methods of ACC data collection in detail, including the setting of mobile devices, the connection between mobile phones and computers, data collection preparation and original data visualization; the fifth section introduces and analyzes UCI data set. Section 6 Segmented ACC data and used Chebyshev high-pass filter to remove the influence of gravity;

In the second chapter :The main purpose of this chapter is to describe the specific algorithm of feature extraction based on time domain and random forest classifier in human activity recognition, so as to obtain better performance.

Time domain features are usually extracted directly from the time series data called "window" using the mean value, standard deviation of activity recognition. The mean is the average of the signal at a particular window and is used to distinguish between static and dynamic activity, such as walking and standing. The stability of the time series signal can be expressed as the standard deviation, and we can use the standard deviation value of the acceleration signal (from the accelerometer) to capture the range of possible acceleration values, which can be used to distinguish the same looking but different speed activities, such as walking and walking up stairs.

Description of random forest classifier, first outlined the random forest algorithm based content, including the generation of decision tree and node splitting algorithm, the generation process of random forest, such as random forest algorithm in two aspects of training and test data of the specific process and CART

algorithm is selected as a random forest decision tree node splitting algorithm, The parameters involved in the tuning include adjusting the number of decision trees in the forest, the maximum depth of decision trees, the maximum number of leaves in decision trees, and so on.

In the third chapter: we set two experiments to evaluate the performance of random forest classifier and compare with KNN, NB, SVM. This experiment suppose HAR system based on random forest can identify six physical human activities and the average accuracy is 94.6%.

In the second experiment-Random forest and multiple classifier comparison experiment, we can get the information：

– TPR- RF (96.1%), SVM (88.1%), KNN (95.74%), NB (81.92%);

– FNR- RF (3.9%), SVM (11.9%), KNN (88.94%), NB (11.06%);

– PPV- RF (96.44%), SVM (88.94%), KNN (95.18%), NB (82.7%);

– FDR- RF (3.56%), SVM (11.06%), KNN (4.82%), NB (17.3%).

This experiment shows four classifier accuracy: RF (96.8%), SVM (88.6%), K-NN (95.3%), and NB (81.1%).The prediction accuracy of the four typical algorithms for human activity recognition, and the final experimental results show that random forests have good performance in human activity recognition.

CONCLUSION

Human activity recognition is one of the most important fields in the real world. In this master thesis, we suppose a human activity recognition system based on random forest which can actually identify six physical human activities, including Walking, Walking downstairs, walking upstairs, Lying, Standing, and Siting.

In this master's thesis, the accelerometer, and gyroscope built into smart phone were used to obtain the original signal, and the sampling rate was set to 50HZ, and the sliding window (128 readings/Windows) with 2.56 seconds and 50% overlap was selected. With time series data as input, the data are preprocessed, feature extraction, model construction and model training on MATLAB.

This thesis uses a random forest classifier for activity recognition to improve the accuracy of recognition. Four typical statistical learning algorithms, including the nearest Neighbor algorithm (KNN), Bayesian algorithm (NB), support vector machine (SVM) and random forest algorithm (RF), were used to construct human activity recognition models for comparison. Classification accuracy was used as the algorithm performance index to judge the recognition accuracy.

Experimental results show that the recognition accuracy of human activities using random forest classifier can reach 96.8%, and the performance of the random

forest algorithm is the best among the SVM (88.6%), K-NN (95.3%), and NB (81.1%).

LIST OF AUTHOR'S PUBLICATIONS

1 Xueying, Y.Human activity recognition based on random forest / Y. Xueying // Proc. of International Scientific and Technical Seminar, Technologies for the transmission and processing of information, Minsk. –2022. – P.77–81.

2 Xueying,Y.Human activity recognition based on Adaboost ensemble classifier / Y.Xueying, I.Boriskevich// Proc. of International Scientific and Technical Seminar ,Technologies for the transmission and processing of information, Minsk. –2022. – P.25–30.

3 Yang, Xueying.Recognition of human activities based on decision optimization model / Yang, Xueying, Huang Gang // Proc. of 2021 4th International Conference on Algorithms, Computing and Artificial Intelligence. – 2021. – P.1–8.

4 Yang, Xueying.Coordinated path planning for Multi-UAVs based on critical track points / Yang, Xueying,Huang Gang // Proc. of 2021 2nd International Conference on Control, Robotics and Intelligent System. – 2021. – P.48–53.